LEDBAT architecture framework consisting of pluggable components
draft-mayutan-ledbat-congestionarchitecture-00.txt

Abstract

The Low Extra Delay Background Transport (LEDBAT) working group is considering protocols for an alternative congestion control protocol that enables a delay-insensitive networking application to minimize the extra queuing delay it causes to other applications because of additional queueing at the bottleneck, when these connections carrying traffic for such applications attempt to use the available bandwidth.

This document proposes an architectural framework for LEDBAT congestion control mechanisms, based on existing work on congestion control protocols and the requirements of the LEDBAT working group. The architectural framework consists of a LEDBAT-congestion control (LEDBAT-CC) suite that provides flexibility in utilizing different components for providing congestion control for transport connections carrying delay-insensitive traffic. The LEDBAT-CC suite of protocols is envisioned to support the multiple alternative mechanisms for bandwidth estimation, congestion detection and indication and end-system flow control to comprise a network friendly congestion avoidance protocol.

This document is inspired by the need to standardize the various components that constitute the network friendly congestion control protocol to avoid having to individually standardize a multitude of distinct and monolithic solutions.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.
Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on September 3, 2010.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the BSD License.
Table of Contents

1. Introduction .................................................. 4
2. Terminology .................................................. 4
3. Design goals .................................................. 4
4. Architecture framework ........................................ 5
   4.1. Bandwidth-estimation module .............................. 6
       4.1.1. Slow start followed by linear increase ................ 7
       4.1.2. Probing packets based bandwidth estimation .......... 7
       4.1.3. Network support based bandwidth estimation .......... 8
   4.2. Congestion-detection module ............................... 8
       4.2.1. Delay based congestion detection ..................... 8
       4.2.2. Explicit congestion marking based congestion
detection ....................................................... 9
       4.2.3. Loss based congestion detection ....................... 9
   4.3. Flow-control module ...................................... 9
5. Acknowledgements .............................................. 10
6. IANA Considerations ........................................... 10
7. Security Considerations ....................................... 10
8. References .................................................... 10
   8.1. Normative References ..................................... 10
   8.2. Informative References ................................... 10
Authors’ Addresses ................................................. 11
1. Introduction

The Low Extra Delay Background Transport (LEDBAT) working group is considering protocols for an alternative congestion control protocol that can saturate available bandwidth during non-congestion periods and yield to delay sensitive applications that use standard TCP.

While the current LEDBAT proposal [I-D.ietf-ledbat-congestion] is based on delay measurements for congestion detection, it could alternatively use explicit congestion notification (ECN)-based marking [RFC3168] to detect and indicate congestion if the network supports it. Similarly, alternative proposals could be used to enhance its performance to opportunistically use available bandwidth as fast as possible during non-congested scenarios.

This document proposes an architectural framework consisting of a LEDBAT-congestion control (LEDBAT-CC) suite of protocols for a network friendly congestion control protocol. The framework consists of decoupled mechanisms to support available bandwidth estimation, congestion detection and flow control. Such an architecture would make it easier to standardize the various mechanisms instead of having to standardize a multitude of monolithic and independent solutions.

The aim is to both develop and have pluggable components that could be used interchangeably depending on the support provided by the network, the bandwidth delay product (BDP) of the network, and the support provided by the end hosts both at the sender side and the receiver side.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Design goals

The design goals for a network friendly congestion control protocol are listed in the LEDBAT draft [I-D.ietf-ledbat-congestion]. The design goals mention the need for the protocol to be flexible enough to accommodate the use of mechanisms other than those stated in the draft. For e.g the use of explicit congestion notification using ECN-marking [RFC3168].

The LEDBAT-suite architecture framework proposed in this document has
A similar set of design goals as those stated in [I-D.ietf-ledbat-congestion]. Some of the design goals of a network friendly congestion control protocol are:

1: Detect incipient congestion as early as possible

1: Maximize utilization of the available bandwidth when the network is not congested

2: Maintain a low delay at the buffers when no other traffic is present

3: Be fair to other network friendly approaches as long as there is no congestion

4: Quickly yield to traffic sharing the same bottleneck queue that uses standard TCP congestion control

5: Add little to the queuing delays induced by TCP traffic

6: Operate well in different kinds of networks, and where applicable use network support

7: Be deployable for delay-insensitive applications that currently comprise noticeable fractions of Internet traffic

4. Architecture framework

Among its other tasks, TCP controls bandwidth estimation, end-system dynamic flow control and reacting to indications of congestion that are based on either end-systems or the network detecting congestion. A network friendly TCP would need to detect the availability of bandwidth as quickly as possible to enable it to opportunistically utilize the available bandwidth. Moreover it needs to be backed up with an efficient congestion detection mechanism and a robust reaction mechanism to incipient congestion.

This section describes the overall architectural framework consisting of distinct component mechanisms that contribute to the proper functioning of a network friendly congestion control protocol. The architecture consists of decoupled mechanisms and we provide some examples of the various pluggable components that could be used for each mechanism. Figure 1 shows the types of pluggable components that can be used as each distinct component.
4.1. Bandwidth-estimation module

Bandwidth estimation is the mechanism used by TCP to try to estimate the available bandwidth. Standard TCP schemes like RENO identify available bandwidth by placing a load on the network and exploring to determine the available bandwidth of the network. Standard TCP has been shown to be insufficient in high bandwidth delay product (BDP) networks, because of its conservative approach to increase the load on the network. Other schemes such as HighSpeed TCP [RFC3649], CUBIC TCP [I-D.rhee-tcpm-cubic] and Quick-Start for TCP [RFC4782] try to capture the bandwidth more quickly.
An efficient bandwidth utilization mechanism backed by an efficient
detection of incipient congestion will potentially improve the
throughput of a network friendly TCP protocol during non-congestion
periods. Since the network friendly congestion control protocol
desires to be "submissive" in the presence of standard TCP during
network congestion, an efficient bandwidth estimation scheme will
enable it to be opportunistic in using available bandwidth and thus
optimize the throughput it achieves as well as increase network
utilization without causing congestion. The choice of a good
starting behaviour could also depend on the network topology, the BDP
of the network, and the support provided by the network.

A Bandwidth-estimation module would be expected to provide a
reasonable estimate of the available bandwidth (BW_ESTIMATE) to the
Flow-control module. The subsequent subsections introduce some of
the prevalent bandwidth estimation schemes. A more exhaustive list
will be added in the follow up version of this document.

4.1.1. Slow start followed by linear increase

This is the standard TCP’s implicit bandwidth estimation mechanism
that works by placing load on the network until it causes congestion
(queueing or loss at a bottleneck in the path). It starts with a
slow start phase where it places an extra load of one packet per
acknowledgement received and switches over to a linear increase phase
wherein it places a load of an extra packet per RTT.

4.1.2. Probing packets based bandwidth estimation

The standard TCP bandwidth estimation mechanism has been shown to be
insufficient in a high BDP network. An alternative bandwidth
estimation method could be based on probing for available bandwidth.

For example, in a pathchirp-like probing mechanism, the end host
sends a train of packets interspaced by specified intervals and tries
to probe for an estimate of available bandwidth between a lowrate and
highrate. The estimate is driven off of changes in the delays
experienced by the packets, and works in a somewhat similar
conceptual approach as that used in delay-based methods for
congestion control. Except here, we do not intimately tie the flow
control reaction to the delay experienced by the packets. The range
depends on various factors such as the number of probe packets, the
spread factor etc. The choice of lowrate and highrate can be varied
to perform dynamic bandwidth estimation.
4.1.3. Network support based bandwidth estimation

Another means to overcome some of the limitations of slow start is to request the routers along the path to help with the bandwidth estimation.

For example, in a Quick-Start TCP [RFC4782] like mechanism, the end host would indicate its desired sending rate using a quick start option in the IP header of the TCP packet. Each router along the path will in turn, either approve the rate or reduce the rate. The receiver would in turn indicate the final rate to the sender. The router should be aware of the Quick-Start option to be able to help.

4.2. Congestion-detection module

A network friendly congestion control avoidance protocol must be backed by an efficient congestion detection mechanism that is able to detect incipient congestion earlier than standard TCP, to enable it to react earlier. In the case of a network friendly congestion control avoidance protocol, it is all the more important that the protocol is be able to detect incipient congestion earlier than standard TCP is able to do so, so especially the loss-based congestion detection). Early detection of congestion assists a network friendly congestion control avoidance protocol in quickly yielding to standard TCP.

The aim is to ensure that the packets of network friendly application packets do not contribute to queue build-up/congestion that results in higher latencies for delay-sensitive traffic and other traffic that may use TCP. The congestion detection scheme needs to ensure that the protocol is able to detect the onset congestion as early as reasonable, while ensuring that it doesn’t respond to truly short-term transients.

The congestion-detection module is expected to send provide a reliable CONGESTION_DETECTED variable indication to the flow-control module. Some example congestion detection mechanisms currently available are listed in the subsequent subsections.

4.2.1. Delay based congestion detection

Delay based schemes for detection of incipient congestion depend on the difference in the current one-way delay to that of the observed base delay. The advantage of such a scheme is that it does not need network support and is easy to deploy. The current LEDBAT effort is based on such a mechanism to detect congestion. However, there is the potential that it may not be an accurate indicator of congestion, especially if the base delay was based on a congested scenario or if
there are frequent route changes.

4.2.2. Explicit congestion marking based congestion detection

A marking based scheme could depend on receiving explicit congestion marking from congested routers. Such a scheme would require the support of intermediate network nodes. This scheme could make use of the existing congestion marking schemes such as ECN marking [RFC3168], PCN marking [RFC5559] using the existing RED queue mechanisms. Additionally, it may be feasible to propose additional, new marking behaviours.

4.2.3. Loss based congestion detection

For completion, we mention a loss based congestion detection mechanism. The network friendly congestion control protocol may use packet loss as an indicator to detect congestion. But for this scheme to be effective, the network should be able to induce drops for the network friendly applications earlier than it does for standard TCP.

4.3. Flow-control module

The flow-control module is in charge of the sending rate. Based on the feedback received from the bandwidth-estimation module and the congestion-detection module, it either increases it’s sending rate or decreases it.

The sending rate can be based on the standard TCP like mechanism, wherein slow start is followed by a linear increase. To optimize the network throughput and be opportunistic in using available bandwidth, a network friendly congestion control mechanism may be designed to be aggressive during non-congestion periods. During this phase, the protocol can use the estimated bandwidth as a means to either perform multiplicative increase or jump directly to the estimated value. If the protocol is backed by an efficient congestion detection and reaction to congestion scheme, it can attempt to aggressively utilize available bandwidth. This is driven by the consideration that not using available bandwidth means that it is most likely left underutilized.

The decrease rate on detection of congestion is an important feature of a network friendly congestion avoidance protocol. It must be able to yield as quickly as possible to standard TCP on the detection of incipient congestion. Standard TCP reduces its congestion window by half. But we may envisage schemes that have a more “severe” multiplicative decrease or even a drastic reduction by reducing the congestion window to one in an attempt to be extremely network
friendly. The reduction rate could also depend on the accuracy of the congestion detection and available bandwidth estimation mechanisms. Further investigation is necessary to arrive at a good decrease rate on the detection of incipient congestion, depending on the means used for detection as well as the bandwidth estimation method.

5. Acknowledgements

6. IANA Considerations

None at this time

7. Security Considerations

None at this time

8. References

8.1. Normative References


8.2. Informative References


[RFC4782]  Floyd, S., Allman, M., Jain, A., and P. Sarolahti, "Quick-


Authors' Addresses

Mayutan Arumaithurai (editor)
University of Goettingen
Email: arumaithurai@cs.uni-goettingen.de

Xiaoming Fu
University of Goettingen
Email: fu@cs.uni-goettingen.de

K. K. Ramakrishnan
AT&T Labs-Research
Email: kkrama@research.att.com